NONCOMMUTATIVE HARMONIC ANALYSIS
WITH APPLICATIONS TO PROBABILITY II
BANACH CENTER PUBLICATIONS, VOLUME 89
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2010

A NEW APPROACH TO MUTUAL INFORMATION. II

FUMIO HIAI and TAKUHO MIYAMOTO

Graduate School of Information Sciences, Tohoku University
Aoba-ku, Sendai 980-8579, Japan
E-mail: hiai@math.is.tohoku.ac.jp, miya@ims.is.tohoku.ac.jp

Abstract. A new concept of mutual pressure is introduced for potential functions on both continuous and discrete compound spaces via discrete micro-states of permutations, and its relations with the usual pressure and the mutual information are established. This paper is a continuation of the paper of Hiai and Petz in Banach Center Publications, Vol. 78.

Introduction. Entropy and pressure are two basic quantities in statistical physics as well as information theory, which are in the duality relation via the Legendre transforms of each other. Mutual information is another important entropic quantity in information theory. The aim of this paper is to seek for the mutual version of pressure whose Legendre transform is equal to the mutual information.

The mutual information of two random variables X and Y is defined as the relative entropy

$$I(X \wedge Y) := S(\mu_{(X,Y)} \,\|\, \mu_X \otimes \mu_Y),$$

where $\mu_{(X,Y)}$ is the joint distribution measure of (X,Y) and $\mu_X \otimes \mu_Y$ is the product of the respective distribution measures of X, Y. This is also expressed as

$$I(X \wedge Y) = -S(X,Y) + S(X) + S(Y)$$

in terms of the Shannon entropy $S(\cdot)$ when X, Y are discrete random variables. When X, Y are continuous variables, the expression holds with the Boltzmann-Gibbs entropy $H(\cdot)$ in place of $S(\cdot)$ (as long as $H(X)$ and $H(Y)$ are finite). These definitions and expressions are naturally extended to the case of more than two random variables.

In the classical (= commutative) probability setting, we developed in the previous paper [5] a certain "discretization approach" to the mutual information by using "discrete

2000 *Mathematics Subject Classification*: Primary 62B10, 94A17.
Key words and phrases: mutual information, pressure, mutual pressure, Boltzmann-Gibbs entropy, Shannon entropy, relative entropy, micro-state, type, Sanov theorem.
The paper is in final form and no version of it will be published elsewhere.

DOI: 10.4064/bc89-0-9

micro-states" of permutations. In this paper we apply the same idea to introduce the notion of the "mutual pressure" for (continuous) potential functions on compound phase spaces. We consider the n-fold product $[-R,R]^n$ of the bounded interval $[-R,R]$, which is regarded as the phase space for an n-tuple of real bounded random variables. For a real continuous function h on $[-R,R]^n$ the usual pressure of h is given by

$$P(h) := \log \int_{[-R,R]^n} e^{h(\mathbf{x})}\,d\mathbf{x}.$$

For an n-tuple $(\mu_1,\dots,\mu_n)$ of probability measures on $[-R,R]$, we choose an approximating sequence $(\xi_1(N),\dots,\xi_n(N))$ such that $\xi_i(N)$ are vectors in $[-R,R]^N_{\le}$ (having the coordinates in increasing order) and $\xi_i(N)\to\mu_i$ in moments as $N\to\infty$ for $1\le i\le n$. We define the mutual pressure $P_{\mathrm{sym}}(h:\mu_1,\dots,\mu_n)$ of h with respect to $(\mu_1,\dots,\mu_n)$ to be the lim sup as $N\to\infty$ of the asymptotic average

$$\frac{1}{N}\log\biggl[\frac{1}{(N!)^n}\sum_{\sigma_1,\dots,\sigma_n\in S_N}\exp\bigl(N\kappa_N(h(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N)))))\bigr)\biggr],$$

over permutations $\sigma_i\in S_N$, where $\kappa_N(h(\mathbf{x}_1,\dots,\mathbf{x}_n)) := \frac{1}{N}\sum_{j=1}^N h(x_{1j},\dots,x_{nj})$ for $\mathbf{x}_i = (x_{i1},\dots,x_{iN})\in[-R,R]^N$, $1\le i\le n$ (Definition 2.1). Then the inequality

$$P(h)\ge P_{\mathrm{sym}}(h:\mu_1,\dots,\mu_n)+\sum_{i=1}^n H(\mu_i)$$

is shown to hold, and the equality case is characterized in a natural way (Theorem 3.2). Moreover, for a probability measure μ on $[-R,R]^n$ with marginal measures $\mu_1,\dots,\mu_n$ on $[-R,R]$, the Legendre transform of $P_{\mathrm{sym}}(h:\mu_1,\dots,\mu_n)$ is shown to be equal to the mutual information $-H(\mu)+\sum_{i=1}^n H(\mu_i)$ as long as $H(\mu_i)>-\infty$ for $1\le i\le n$ (Theorem 3.5).

The same approach can be also applied to the setting of discrete phase spaces, when the Shannon entropy takes the place of the Boltzmann-Gibbs entropy. We deal with the discrete case in Section 4 separately since the discussions are considerably different from the continuous case due to the difference of entropies.

1. Preliminaries in the continuous case. Let $R>0$ and $n\in\mathbb{N}$ be fixed throughout. We denote by $\mathrm{Prob}([-R,R]^n)$ the set of probability measures on the n-fold product $[-R,R]^n$ ($\subset\mathbb{R}^n$), and by $C_{\mathbb{R}}([-R,R]^n)$ the real Banach space of real continuous functions on $[-R,R]^n$ with the sup-norm $\|f\| := \max\{|f(\mathbf{x})| : \mathbf{x}\in[-R,R]^n\}$. The *Boltzmann-Gibbs entropy* of a probability measure μ on $[-R,R]^n$ is defined to be

$$H(\mu) := -\int_{[-R,R]^n} p(\mathbf{x})\log p(\mathbf{x})\,d\mathbf{x}$$

if μ has the joint density $p(\mathbf{x})$ with respect to the Lebesgue measure $d\mathbf{x}$ on $\mathbb{R}^N$; otherwise $H(\mu):=-\infty$. A measure $\mu\in\mathrm{Prob}([-R,R]^n)$ typically arises as the joint distribution of an n-tuple $(X_1,\dots,X_n)$ of real random variables bounded by R (i.e., $|X_i|\le R$) on a probability space. In this case, we have $H(\mu)=H(X_1,\dots,X_n)$.

For a vector $\mathbf{x} = (x_1, \dots, x_N)$ in $\mathbb{R}^N$ we write $\|\mathbf{x}\|_1 := N^{-1}\sum_{j=1}^N |x_j|$. The mean value of $\mathbf{x}$ is given by

$$\kappa_N(\mathbf{x}) := \frac{1}{N}\sum_{j=1}^N x_j.$$

For each $N, m \in \mathbb{N}$ and $\delta > 0$ we define $\Delta_R(\mu; N, m, \delta)$ to be the set of all n-tuples $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ of $\mathbf{x}_i = (x_{i1}, \dots, x_{iN}) \in [-R, R]^N$, $1 \le i \le n$, such that

$$|\kappa_N(\mathbf{x}_{i_1} \cdots \mathbf{x}_{i_k}) - \mu(x_{i_1} \cdots x_{i_k})| < \delta$$

for all $i_1, \dots, i_k \in \{1, \dots, n\}$ with $1 \le k \le m$, where $\mathbf{x}_{i_1} \cdots \mathbf{x}_{i_k}$ stands for the pointwise product, i.e.,

$$\mathbf{x}_{i_1} \cdots \mathbf{x}_{i_k} := (x_{i_1 1} \cdots x_{i_k 1}, x_{i_1 2} \cdots x_{i_k 2}, \ \dots \ , x_{i_1 N} \cdots x_{i_k N}) \in \mathbb{R}^N,$$

and

$$\mu(x_{i_1} \cdots x_{i_k}) := \int_{[-R,R]^n} x_{i_1} \cdots x_{i_k} \, d\mu(x_1, \dots, x_n).$$

Then it is known [4, 5.1.1] that the limit

$$\lim_{N\to\infty} \frac{1}{N} \log \lambda_N^{\otimes n}(\Delta_R(\mu; N, m, \delta))$$

exists, where λ_N stands for the Lebesgue measure on $\mathbb{R}^N$, and furthermore we have

$$H(\mu) = \lim_{m\to\infty, \delta\searrow 0} \lim_{N\to\infty} \frac{1}{N} \log \lambda_N^{\otimes n}(\Delta_R(\mu; N, m, \delta)).$$

In [5] we introduced some kinds of mutual information $I_{\mathrm{sym}}(\mu)$ and $\overline{I}_{\mathrm{sym}}(\mu)$, and established their relations with $H(\mu)$ as follows.

Definition 1.1. Let $\mu \in \mathrm{Prob}([-R, R]^n)$ and μ_i be the restriction (or the marginal) of μ to the ith component $[-R, R]$ of $[-R, R]^n$ for $1 \le i \le n$. Choose and fix a sequence of n-tuples $\Xi(N) = (\xi_1(N), \dots, \xi_n(N))$, $N \in \mathbb{N}$, of $\mathbb{R}^N$-vectors $\xi_i(N)$ in $[-R, R]^N_{\le} := \{(x_1, \dots, x_N) \in [-R, R]^N : x_1 \le \dots \le x_N\}$ such that $\kappa_N(\xi_i(N)^k) \to \int x^k \, d\mu_i(x)$ as $N \to \infty$ for all $k \in \mathbb{N}$, i.e., $\xi_i(N) \to \mu_i$ in moments for $1 \le i \le n$. We call such a sequence $\Xi(N)$ an *approximating sequence* for $(\mu_1, \dots, \mu_n)$. For $N \in \mathbb{N}$ the action of the symmetric group S_N on $\mathbb{R}^N$ is given by

$$\sigma(\mathbf{x}) := (x_{\sigma^{-1}(1)}, \dots, x_{\sigma^{-1}(N)})$$

for $\sigma \in S_N$ and $\mathbf{x} = (x_1, \dots, x_N) \in \mathbb{R}^N$. For each $N, m \in \mathbb{N}$ and $\delta > 0$ we define $\Delta_{\mathrm{sym}}(\mu : \Xi(N); N, m, \delta)$ to be the set of all $(\sigma_1, \dots, \sigma_n) \in S_N^n$ such that

$$(\sigma_1(\xi_1(N)), \dots, \sigma_n(\xi_n(N))) \in \Delta_R(\mu; N, m, \delta).$$

We define

$$I_{\mathrm{sym}}(\mu) := -\lim_{m\to\infty, \delta\searrow 0} \limsup_{N\to\infty} \frac{1}{N} \log \gamma_{S_N}^{\otimes n}(\Delta_{\mathrm{sym}}(\mu : \Xi(N); N, m, \delta)),$$

where γ_{S_N} is the uniform probability measure on S_N, and define also $\overline{I}_{\mathrm{sym}}(\mu)$ by replacing lim sup by lim inf. This definitions of $I_{\mathrm{sym}}(\mu)$ and $\overline{I}_{\mathrm{sym}}(\mu)$ are independent of the choice of an approximating sequence $\Xi(N)$ for $(\mu_1, \dots, \mu_n)$ ([5, Lemma 1.5]).

THEOREM 1.2 ([5, Theorem 1.6]). *For every* $\mu \in \mathrm{Prob}([-R,R]^n)$ *with marginals* $\mu_1,\dots,\mu_n \in \mathrm{Prob}([-R,R])$,

$$H(\mu) = -I_{\mathrm{sym}}(\mu) + \sum_{i=1}^n H(\mu_i) = -\overline{I}_{\mathrm{sym}}(\mu) + \sum_{i=1}^n H(\mu_i).$$

The *pressure* of $h \in C_{\mathbb{R}}([-R,R]^n)$ is given by

$$P(h) := \log \int_{[-R,R]^n} e^{h(\mathbf{x})}\,d\mathbf{x}.$$

It is well known that the pressure function $P(h)$ for $h \in C_{\mathbb{R}}([-R,R]^n)$ and the (minus) Boltzmann-Gibbs entropy $-H(\mu)$ for $\mu \in \mathrm{Prob}([-R,R]^n)$ are in the duality relation in the sense that they are the Legendre transforms of each other. That is,

$$\begin{aligned} H(\mu) &= \inf\{-\mu(h) + P(h) : h \in C_{\mathbb{R}}([-R,R]^n)\}, \quad \mu \in \mathrm{Prob}([-R,R]^n), \\ P(h) &= \max\{\mu(h) + H(\mu) : \mu \in \mathrm{Prob}([-R,R]^n)\}, \quad h \in C_{\mathbb{R}}([-R,R]^n). \end{aligned} \tag{1.1}$$

Furthermore, for every $h \in C_{\mathbb{R}}([-R,R]^n)$ the *Gibbs probability measure* μ_h associated with h is given by

$$d\mu_h(\mathbf{x}) := \frac{1}{Z_h} e^{h(\mathbf{x})}\,d\mathbf{x} \quad \text{with} \quad Z_h := \int_{[-R,R]^n} e^{h(\mathbf{x})}\,d\mathbf{x} = e^{P(h)},$$

which is characterized by the variational equality

$$P(h) = \mu_h(h) + H(\mu_h),$$

that is, μ_h is a unique maximizer of $\mu \in \mathrm{Prob}([-R,R]^n) \mapsto \mu(h) + H(\mu)$.

2. Mutual pressure and its Legendre transform. In the setting of continuous compound spaces described in Section 1, we introduce the mutual version of pressure for continuous potential functions, and consider its Legendre transform that is a version of the mutual information.

DEFINITION 2.1. Let $\mu_1,\dots,\mu_n \in \mathrm{Prob}([-R,R])$ be given and choose an approximating sequence $\Xi(N) = (\xi_1(N),\dots,\xi_n(N))$ of $\xi_i(N) \in [-R,R]^N_\le$ for $(\mu_1,\dots,\mu_n)$ as in Definition 1.1. For each $h \in C_{\mathbb{R}}([-R,R]^n)$ and $\mathbf{x}_i = (x_{i1},\dots,x_{iN}) \in [-R,R]^N$, $1 \le i \le n$, define

$$h(\mathbf{x}_1,\dots,\mathbf{x}_n) := (h(x_{11},\dots,x_{n1}), h(x_{12},\dots,x_{n2}),\dots,h(x_{1N},\dots,x_{nN})) \in \mathbb{R}^N \tag{2.1}$$

and hence

$$\kappa_N(h(\mathbf{x}_1,\dots,\mathbf{x}_n)) := \frac{1}{N}\sum_{j=1}^N h(x_{1j},\dots,x_{nj}). \tag{2.2}$$

For each $h \in C_{\mathbb{R}}([-R,R]^n)$ we define the *mutual pressure* of h with respect to $(\mu_1,\dots,\mu_n)$ to be

$$\begin{aligned}
&P_{\mathrm{sym}}(h:\mu_1,\dots,\mu_n)\\
&\quad:=\limsup_{N\to\infty}\frac{1}{N}\log\int_{S_N^n}\exp\bigl(N\kappa_N(h(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N))))\bigr)\,d\gamma_{S_N}^{\otimes n}(\sigma_1,\dots,\sigma_n)\\
&\quad=\limsup_{N\to\infty}\frac{1}{N}\log\biggl[\frac{1}{(N!)^n}\sum_{\sigma_1,\dots,\sigma_n\in S_N}\exp\bigl(N\kappa_N(h(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N))))\bigr)\biggr].
\end{aligned}$$

The above definition is justified by the following:

LEMMA 2.2. *$P_{\mathrm{sym}}(h:\mu_1,\dots,\mu_n)$ is independent of the choice of an approximating sequence $\Xi(N)$ for $(\mu_1,\dots,\mu_n)$.*

Proof. Let $\Xi'(N)=(\xi_1'(N),\dots,\xi_n'(N))$ be another such approximating sequence. We write $P_{\mathrm{sym}}(h:\Xi)$ and $P_{\mathrm{sym}}(h:\Xi')$ for $P_{\mathrm{sym}}(h:\mu_1,\dots,\mu_n)$ defined in Definition 2.1 with $\Xi(N)$ and $\Xi'(N)$, respectively. Since $P_{\mathrm{sym}}(h:\Xi)$ and $P_{\mathrm{sym}}(h:\Xi')$ are continuous in h in the norm (see Proposition 2.3 (3) below), it suffices to prove that $P_{\mathrm{sym}}(p:\Xi)=P_{\mathrm{sym}}(p:\Xi')$ for any real polynomial p of n variables $x_1,\dots,x_n$. Since $\xi_i(N),\xi_i'(N)\in[-R,R]^N_{\le}$, for any $\varepsilon>0$ there exists a $\delta>0$ such that, for every $N\in\mathbb{N}$, if $\|\xi_i(N)-\xi_i'(N)\|_1<\delta$ for all $i=1,\dots,n$, then

$$|\kappa_N(p(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N))))-\kappa_N(p(\sigma_1(\xi_1'(N)),\dots,\sigma_n(\xi_n'(N))))|<\varepsilon$$

for all $(\sigma_1,\dots,\sigma_n)\in S_N^n$. Thanks to [6, Lemma 4.3] (also [4, 4.3.4]), there exists an $N_0\in\mathbb{N}$ such that if $N\ge N_0$ then $\|\xi_i(N)-\xi_i'(N)\|_1<\delta$ for all $i=1,\dots,n$. Hence we have for every $N\ge N_0$

$$\begin{aligned}
&\biggl|\frac{1}{N}\log\biggl[\frac{1}{(N!)^n}\sum_{\sigma_1,\dots,\sigma_n\in S_N}\exp\bigl(N\kappa_N(p(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N))))\bigr)\biggr]\\
&\qquad-\frac{1}{N}\log\biggl[\frac{1}{(N!)^n}\sum_{\sigma_1,\dots,\sigma_n\in S_N}\exp\bigl(N\kappa_N(p(\sigma_1(\xi_1'(N)),\dots,\sigma_n(\xi_n'(N))))\bigr)\biggr]\biggr|<\varepsilon.
\end{aligned}$$

This implies that $|P_{\mathrm{sym}}(p:\Xi)-P_{\mathrm{sym}}(p:\Xi')|\le\varepsilon$. Since $\varepsilon>0$ is arbitrary, the desired conclusion follows. ■

The following are basic properties of $P_{\mathrm{sym}}(h:\mu_1,\dots,\mu_n)$, whose proofs are straightforward.

PROPOSITION 2.3. *Let $\mu_1,\dots,\mu_n\in\mathrm{Prob}([-R,R])$.*

(1) *When $n=1$, $P_{\mathrm{sym}}(h:\mu_1)=\mu_1(h)$ for all $h\in C_{\mathbb{R}}([-R,R])$.*
(2) *$P_{\mathrm{sym}}(h:\mu_1,\dots,\mu_n)$ is a convex and increasing function on $C_{\mathbb{R}}([-R,R]^n)$.*
(3) *$|P_{\mathrm{sym}}(h:\mu_1,\dots,\mu_n)-P_{\mathrm{sym}}(h':\mu_1,\dots,\mu_n)|\le\|h-h'\|$ for all $h,h'\in C_{\mathbb{R}}([-R,R]^n)$.*
(4) *If $1\le m<n$, $h^{(1)}\in C_{\mathbb{R}}([-R,R]^m)$, $h^{(2)}\in C_{\mathbb{R}}([-R,R]^{n-m})$ and $h(x_1,\dots,x_n):=h^{(1)}(x_1,\dots,x_m)+h^{(2)}(x_{m+1},\dots,x_n)$, then*

$$P_{\mathrm{sym}}(h:\mu_1,\dots,\mu_n)\le P_{\mathrm{sym}}(h^{(1)}:\mu_1,\dots,\mu_m)+P_{\mathrm{sym}}(h^{(2)}:\mu_{m+1},\dots,\mu_n).$$

DEFINITION 2.4. Let $\mu\in\mathrm{Prob}([-R,R]^n)$ with marginals $\mu_1,\dots,\mu_n\in\mathrm{Prob}([-R,R])$. Define

$$\mathcal{I}_{\mathrm{sym}}(\mu):=\sup\{\mu(h)-P_{\mathrm{sym}}(h:\mu_1,\dots,\mu_n):h\in C_{\mathbb{R}}([-R,R]^n)\},$$

that is, $\mathcal{I}_{\rm sym}(\mu)$ is the Legendre transform of $P_{\rm sym}(h:\mu_1,\dots,\mu_n)$. Furthermore, we say that μ is *mutually equilibrium* associated with $h\in C_{\mathbb{R}}([-R,R]^n)$ if the variational equality

$$\mathcal{I}_{\rm sym}(\mu)=\mu(h)-P_{\rm sym}(h:\mu_1,\dots,\mu_n)$$

holds.

The next proposition says that $P_{\rm sym}(h:\mu_1,\dots,\mu_n)$ is the converse Legendre transform of $\mathcal{I}_{\rm sym}(\mu)$.

Proposition 2.5. *For every $h\in C_{\mathbb{R}}([-R,R]^n)$ and $\mu_1,\dots,\mu_n\in{\rm Prob}([-R,R])$,*

$$P_{\rm sym}(h:\mu_1,\dots,\mu_n)=\max\{\mu(h)-\mathcal{I}_{\rm sym}(\mu):\mu\in{\rm Prob}_{\mu_1,\dots,\mu_n}([-R,R]^n)\},$$

where ${\rm Prob}_{\mu_1,\dots,\mu_n}([-R,R]^n)$ is the set of all $\mu\in{\rm Prob}([-R,R]^n)$ whose restriction to the ith component of $[-R,R]^n$ is μ_i for $1\le i\le n$. Hence there exists a mutually equilibrium probability measure associated with h whose marginals are $\mu_1,\dots,\mu_n$.

Proof. One can consider ${\rm Prob}([-R,R]^n)$ as a closed convex subset of the dual (real) Banach space $C_{\mathbb{R}}([-R,R]^n)^*$ of $C_{\mathbb{R}}([-R,R]^n)$. Let $F:C_{\mathbb{R}}([-R,R]^n)^*\to(-\infty,+\infty]$ be the conjugate (or the Legendre transform) of $P_{\rm sym}(h:\mu_1,\dots,\mu_n)$, i.e.,

$$F(\psi):=\sup\{\psi(h)-P_{\rm sym}(h:\mu_1,\dots,\mu_n):h\in C_{\mathbb{R}}([-R,R]^n)\}$$

for $\psi\in C_{\mathbb{R}}([-R,R]^n)^*$. We then prove that

$$\begin{cases} F(\mu)=\mathcal{I}_{\rm sym}(\mu) & \text{if } \mu\in{\rm Prob}_{\mu_1,\dots,\mu_n}([-R,R]^n),\\ F(\psi)=+\infty & \text{if } \psi\in C_{\mathbb{R}}([-R,R]^n)^*\setminus{\rm Prob}_{\mu_1,\dots,\mu_n}([-R,R]^n).\end{cases} \tag{2.3}$$

The first equality is just the definition of $\mathcal{I}_{\rm sym}(\mu)$. The second follows from the following three claims.

(a) If $\psi(h)<0$ for some $h\in C_{\mathbb{R}}([-R,R]^n)$ with $h\ge0$, then $F(\psi)=+\infty$. In fact, for $\alpha<0$ we have $P_{\rm sym}(\alpha h:\mu_1,\dots,\mu_n)\le P_{\rm sym}(0:\mu_1,\dots,\mu_n)=0$ by Proposition 2.3 (2) so that

$$\psi(\alpha h)-P_{\rm sym}(\alpha h:\mu_1,\dots,\mu_n)\ge\alpha\psi(h)\to+\infty$$

as $\alpha\to-\infty$.

(b) If $\psi(\mathbf{1})\ne1$, then $F(\psi)=+\infty$. In fact, since $P_{\rm sym}(\alpha\mathbf{1}:\mu_1,\dots,\mu_n)=\alpha$ for $\alpha\in\mathbb{R}$, it follows that

$$\psi(\alpha h)-P_{\rm sym}(\alpha\mathbf{1}:\mu_1,\dots,\mu_n)=\alpha(\psi(\mathbf{1})-1)\to+\infty$$

as $\alpha\to+\infty$ or $-\infty$ accordingly as $\psi(\mathbf{1})<1$ or $\psi(\mathbf{1})>1$.

(c) Assume that $\mu\in{\rm Prob}([-R,R]^n)$ but $\mu\notin{\rm Prob}_{\mu_1,\dots,\mu_n}([-R,R]^n)$. Then there exists an $f\in C_{\mathbb{R}}([-R,R])$ such that $\mu(f^{(i)})>\mu_i(f)$ for some $1\le i\le n$, where $f^{(i)}(\mathbf{x}):=f(x_i)$ for $\mathbf{x}=(x_1,\dots,x_n)\in[-R,R]^n$. Since

$$P_{\rm sym}(\alpha f^{(i)}:\mu_1,\dots,\mu_n)=\lim_{N\to\infty}\alpha f(\xi_i(N))=\alpha\mu_i(f)$$

for $\alpha\in\mathbb{R}$, it follows that

$$\mu(\alpha f^{(i)})-P_{\rm sym}(\alpha f^{(i)}:\mu_1,\dots,\mu_n)=\alpha(\mu(f^{(i)})-\mu_i(f))\to+\infty$$

as $\alpha\to+\infty$.

Hence (2.3) is proved. Since $P_{\rm sym}(h : \mu_1, \dots, \mu_n)$ is a convex continuous function on $C_{\mathbb R}([-R,R]^n)$ by Proposition 2.3, the duality theorem for conjugate functions implies that

$$\begin{aligned} P_{\rm sym}(h : \mu_1, \dots, \mu_n) &= \sup\{\psi(h) - F(\psi) : \psi \in C_{\mathbb R}([-R,R]^n)^*\} \\ &= \sup\{\mu(h) - \mathcal{I}_{\rm sym}(\mu) : \mu \in {\rm Prob}_{\mu_1,\dots,\mu_n}([-R,R]^n)\}. \end{aligned}$$

Since ${\rm Prob}_{\mu_1,\dots,\mu_n}([-R,R]^n)$ is weakly* compact and $\mathcal{I}_{\rm sym}(\mu)$ is weakly* lower semicontinuous on ${\rm Prob}_{\mu_1,\dots,\mu_n}([-R,R]^n)$, the above latter supremum is attained by some $\mu \in {\rm Prob}_{\mu_1,\dots,\mu_n}([-R,R]^n)$. ■

PROPOSITION 2.6. *The function* $P_{\rm sym}(h : \mu_1, \dots, \mu_n)$ *is jointly continuous on* $C_{\mathbb R}([-R,R]^n) \times ({\rm Prob}([-R,R]))^n$ *with respect to the norm topology on* $C_{\mathbb R}([-R,R]^n)$ *and the weak* topology on* ${\rm Prob}([-R,R])$.

Proof. Let $h, h' \in C_{\mathbb R}([-R,R]^n)$ and $\mu_i, \mu_i' \in {\rm Prob}([-R,R])$, $1 \le i \le n$. For any $\varepsilon > 0$ choose a real polynomial p of n variables $x_1, \dots, x_n$ such that $\|p - h\| < \varepsilon$. We have

$$\begin{aligned} &|P_{\rm sym}(h : \mu_1, \dots, \mu_n) - P_{\rm sym}(h' : \mu_1', \dots, \mu_n')| \\ &\quad \le |P_{\rm sym}(h : \mu_1, \dots, \mu_n) - P_{\rm sym}(p : \mu_1, \dots, \mu_n)| \\ &\qquad + |P_{\rm sym}(p : \mu_1, \dots, \mu_n) - P_{\rm sym}(p : \mu_1', \dots, \mu_n')| \\ &\qquad + |P_{\rm sym}(p : \mu_1', \dots, \mu_n') - P_{\rm sym}(h' : \mu_1', \dots, \mu_n')| \\ &\quad \le \|h - p\| + \|p - h'\| + |P_{\rm sym}(p : \mu_1, \dots, \mu_n) - P_{\rm sym}(p : \mu_1', \dots, \mu_n')| \\ &\quad \le 2\varepsilon + \|h - h'\| + |P_{\rm sym}(p : \mu_1, \dots, \mu_n) - P_{\rm sym}(p : \mu_1', \dots, \mu_n')| \end{aligned}$$

by Proposition 2.3 (3). Recall that the weak* topology on ${\rm Prob}([-R,R])$ is metrizable with the metric $\rho(\nu,\nu') := \sum_{k=1}^\infty (2R)^{-k}|\nu(x^k) - \nu'(x^k)|$, where $\nu(x^k) := \int x^k \, d\nu(x)$. It suffices to show that there exists a $\delta > 0$ such that if $\rho(\mu_i, \mu_i') < \delta$ for $1 \le i \le n$, then

$$|P_{\rm sym}(p : \mu_1, \dots, \mu_n) - P_{\rm sym}(p : \mu_1', \dots, \mu_n')| \le \varepsilon.$$

One can choose a $\delta_1 > 0$ such that, for every $N \in \mathbb N$, if $\mathbf{x}_i, \mathbf{x}_i' \in [-R,R]^N_\le$ and $\|\mathbf{x}_i - \mathbf{x}_i'\|_1 < \delta_1$ for $1 \le i \le n$, then

$$|\kappa_N(p(\sigma_1(\mathbf{x}_1), \dots, \sigma_n(\mathbf{x}_n))) - \kappa_N(p(\sigma_1(\mathbf{x}_1'), \dots, \sigma_n(\mathbf{x}_n')))| < \varepsilon$$

for all $(\sigma_1, \dots, \sigma_n) \in S_N^n$. Thanks to [6, Lemma 4.3] one can choose an $m \in \mathbb N$ and a $\delta_2 > 0$ such that, for every $N \in \mathbb N$, if $\mathbf{x}, \mathbf{x}' \in [-R,R]^N_\le$ and $|\kappa_N(\mathbf{x}^k) - \kappa_N(\mathbf{x}'^k)| < \delta_2$ for all $k = 1, \dots, m$, then $\|\mathbf{x} - \mathbf{x}'\|_1 < \delta_1$. Then choose a $\delta_3 > 0$ such that if $\nu, \nu' \in {\rm Prob}([-R,R])$ and $\rho(\nu,\nu') < \delta_3$, then $|\nu(x^k) - \nu'(x^k)| < \delta_2/2$ for all $k = 1, \dots, m$. Now assume that $\mu_i, \mu_i' \in {\rm Prob}([-R,R])$ and $\rho(\mu_i, \mu_i') < \delta_3$ for $1 \le i \le n$. Let $\Xi(N) = (\xi_1(N), \dots, \xi_n(N))$ and $\Xi'(N) = (\xi_1'(N), \dots, \xi_n'(N))$ be approximating sequences for $(\mu_1, \dots, \mu_n)$ and $(\mu_1', \dots, \mu_n')$, respectively, with $\xi_i(N), \xi_i'(N) \in [-R,R]^N_\le$. There exists an $N_0 \in \mathbb N$ such that if $N \ge N_0$ then for $1 \le i \le n$ we have

$$\begin{aligned} &|\kappa_N(\xi_i(N)^k) - \kappa_N(\xi_i'(N)^k)| \\ &\quad \le |\kappa_N(\xi_i(N)^k) - \mu_i(x^k)| + |\mu_i(x^k) - \mu_i'(x^k)| + |\mu_i'(x^k) - \kappa_n(\xi_i'(N)^k)| < \delta_2 \end{aligned}$$

for all $k = 1, \dots, m$ so that $\|\xi_i(N) - \xi_i'(N)\|_1 < \delta_1$. Hence if $N \ge N_0$ then we have

$$|\kappa_N(p(\sigma_1(\xi_1(N)), \dots, \sigma_n(\xi_n(N)))) - \kappa_N(p(\sigma_1(\xi_1'(N)), \dots, \sigma_n(\xi_n'(N))))| < \varepsilon$$

for all $(\sigma_1,\dots,\sigma_n)\in S_N^n$. This implies that $|P_{\rm sym}(p:\mu_1,\dots,\mu_n)-P_{\rm sym}(p:\mu_1',\dots,\mu_n')|\le\varepsilon$, as required. ∎

COROLLARY 2.7. *$\mathcal{I}_{\rm sym}(\mu)$ is weakly* lower semicontinuous on* Prob$([-R,R]^n)$.

Proof. Let μ and $\mu^{(k)}$, $k\in\mathbb{N}$, be in Prob$([-R,R]^n)$ such that $\mu^{(k)}\to\mu$ weakly*. Let μ_i and $\mu_i^{(k)}$, $1\le i\le n$, be the marginals of μ and $\mu^{(k)}$, respectively. Since $\mu_i^{(k)}\to\mu_i$ weakly* as $k\to\infty$ for $1\le i\le n$, Proposition 2.6 implies that for every $h\in C_{\mathbb{R}}([-R,R]^n)$

$$\begin{aligned}\mu(h)-P_{\rm sym}(h:\mu_1,\dots,\mu_n)&=\lim_{k\to\infty}\{\mu^{(k)}(h)-P_{\rm sym}(h:\mu_1^{(k)},\dots,\mu_n^{(k)})\}\\&\le\liminf_{k\to\infty}\mathcal{I}_{\rm sym}(\mu^{(k)})\end{aligned}$$

so that $\mathcal{I}_{\rm sym}(\mu)\le\liminf_{k\to\infty}\mathcal{I}_{\rm sym}(\mu^{(k)})$, as required. ∎

3. Relations of $P_{\rm sym}(h)$ with $P(h)$ and of $\mathcal{I}_{\rm sym}(\mu)$ with $H(\mu)$. First let us recall the Sanov large deviation in the form suitable for our purpose. Let $h_0\in C_{\mathbb{R}}([-R,R])$ and μ_0 be the Gibbs probability measure associated with h_0, i.e.,

$$d\mu_0(x):=\frac{1}{Z_{h_0}}\,e^{h_0(x)}\,dx\quad\text{with}\quad Z_{h_0}:=\int_{[-R,R]}e^{h_0(x)}\,dx.$$

Consider the infinite product probability space $([-R,R]^\infty,\mu_0^{\otimes\infty})$ and i.i.d. (independent and identically distributed) random variables $x_1,x_2,\dots$ consisting of coordinate variables of $[-R,R]^\infty$. The *Sanov theorem* (see [3, 6.2.10]) says that the empirical measure (random probability measure)

$$\frac{\delta_{x_1}+\cdots+\delta_{x_N}}{N}$$

satisfies the large deviation principle in the scale $1/N$ with the good rate function $S(\mu\,\|\,\mu_0)$ for $\mu\in$ Prob$([-R,R])$, where $S(\mu\,\|\,\mu_0)$ denotes the *relative entropy* (or the *Kullback-Leibler divergence*) of μ with respect to μ_0. That is,

$$\begin{aligned}\limsup_{N\to\infty}\frac{1}{N}\log\mu_0^{\otimes N}\left(\frac{\delta_{x_1}+\cdots+\delta_{x_N}}{N}\in F\right)&\le-\inf\{S(\mu\,\|\,\mu_0):\mu\in F\},\\\liminf_{N\to\infty}\frac{1}{N}\log\mu_0^{\otimes N}\left(\frac{\delta_{x_1}+\cdots+\delta_{x_N}}{N}\in G\right)&\ge-\inf\{S(\mu\,\|\,\mu_0):\mu\in G\}\end{aligned}$$

for every closed subset F and every open subset G of Prob$([-R,R])$ in the weak* topology. As remarked in [4, p. 211], it then follows (based on the Borel-Cantelli lemma) that the empirical measure $(\delta_{x_1}+\cdots+\delta_{x_N})/N$ converges to μ_0 in the weak* topology almost surely. In the next lemma we state some consequences of the above large deviation, which will play a crucial role in our later discussions.

LEMMA 3.1. *Let h_0 and μ_0 be as above. Then:*

(a) For every $m\in\mathbb{N}$ and $\delta>0$,

$$\lim_{N\to\infty}\mu_0^{\otimes N}(\Delta_R(\mu_0;N,m,\delta))=1.$$

(b) If $\mu_1\in$ Prob$([-R,R])$ and $\mu_1\ne\mu_0$, then there exist an $m\in\mathbb{N}$ and a $\delta>0$ such that

$$\limsup_{N\to\infty}\frac{1}{N}\log\mu_0^{\otimes N}(\Delta_R(\mu_1;N,m,\delta))<0.$$

Proof. (a) For each $m \in \mathbb{N}$ and $\delta > 0$ set

$$G(\mu_0; m, \delta) := \{\mu \in \mathrm{Prob}([-R,R]) : |\mu(x^k) - \mu_0(x^k)| < \delta,\ 1 \le k \le m\},$$

which is a weak* neighborhood of μ_0. Note that $\mathbf{x} = (x_1, \ldots, x_N) \in \Delta_R(\mu_0; N, m, \delta)$ is equivalent to $(\delta_{x_1} + \cdots + \delta_{x_N})/N \in G(\mu_0; m, \delta)$. Since $(\delta_{x_1} + \cdots + \delta_{x_N})/N \to \mu_0$ weakly* in the sense of almost sure (with respect to $\mu_0^{\otimes\infty}$) as remarked above, we have

$$\mu_0^{\otimes N}(\Delta_R(\mu_0; N, m, \delta)) = \mu_0^{\otimes N}\biggl(\frac{\delta_{x_1} + \cdots + \delta_{x_N}}{N} \in G(\mu_0; m, \delta)\biggr) \to 1$$

as $N \to \infty$.

(b) Let $\mu_1 \in \mathrm{Prob}([-R,R])$ with $\mu_1 \ne \mu_0$. One can find an $m \in \mathbb{N}$ and a $\delta > 0$ so that the weak* closed subset

$$F(\mu_1; m, \delta) := \{\mu \in \mathrm{Prob}([-R,R]) : |\mu(t^k) - \mu_1(t^k)| \le \delta,\ 1 \le k \le m\}$$

does not contain μ_0. The large deviation principle implies that

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{1}{N} \log \mu_0^{\otimes N}(\Delta_R(\mu_1; N, m, \delta)) \\
&\quad \le \limsup_{N\to\infty} \frac{1}{N} \log \mu_0^{\otimes N}\biggl(\frac{\delta_{x_1} + \cdots + \delta_{x_N}}{N} \in F(\mu_1; m, \delta)\biggr) \\
&\quad \le -\inf\{S(\mu \,\|\, \mu_0) : \mu \in F(\mu_1; m, \delta)\} < 0,
\end{aligned}$$

because $S(\mu \,\|\, \mu_0)$ is weakly* lower semicontinuous and so attains the minimum (> 0) on a weakly* compact subset $F(\mu_1; m, \delta)$. ∎

The next theorem gives an exact relation between $P_{\mathrm{sym}}(h)$ and $P(h)$.

THEOREM 3.2. *For every $h \in C_{\mathbb{R}}([-R,R]^n)$ and every $\mu_1, \ldots, \mu_n \in \mathrm{Prob}([-R,R])$,*

$$P(h) \ge P_{\mathrm{sym}}(h : \mu_1, \ldots, \mu_n) + \sum_{i=1}^n H(\mu_i). \tag{3.1}$$

Moreover the following conditions are equivalent:

(i) $P(h) = P_{\mathrm{sym}}(h : \mu_1, \ldots, \mu_n) + \sum_{i=1}^n H(\mu_i)$;
(ii) $\mu_1, \ldots, \mu_n$ are the marginals of the Gibbs measure associated with h;
(iii) for each $i = 1, \ldots, n$, μ_i is the Gibbs measure associated with $h_i \in C_{\mathbb{R}}([-R,R])$ defined by

$$h_i(x) := \log \int_{[-R,R]^{n-1}} e^{h(x_1,\ldots,x_{i-1},x,x_{i+1},\ldots,x_n)}\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n$$

for $x \in [-R,R]$.

Proof. Consider the Gibbs probability measure $\mu_h := Z_h^{-1} e^{h(\mathbf{x})}\, d\mathbf{x}$ associated with h so that $P(h) = \log Z_h$. Let $N, m \in \mathbb{N}$ and $\delta > 0$. Then it is straightforward to see that

$$Z_h^N \mu_h^{\otimes N}\Bigl(\prod_{i=1}^n \Delta_R(\mu_i; N, m, \delta)\Bigr) = \int_{\prod_{i=1}^n \Delta_R(\mu_i; N, m, \delta)} \exp\bigl(N\kappa_N(h(\mathbf{x}_1, \ldots, \mathbf{x}_n))\bigr) \prod_{i=1}^n d\mathbf{x}_i,$$

where $\prod_{i=1}^n \Delta_R(\mu_i; N, m, \delta)$ in the left-hand side is regarded as a subset of $(\mathbb{R}^n)^N$ by the correspondence $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \leftrightarrow \bigl((x_{i1})_{i=1}^n, (x_{i2})_{i=1}^n, \ldots, (x_{iN})_{i=1}^n\bigr)$ for $\mathbf{x}_i = (x_{i1}, \ldots, x_{iN})$.

Hence we have

$$
\begin{aligned}
&Z_h^N \mu_h^{\otimes N}\Big(\prod_{i=1}^n \Delta_R(\mu_i;N,m,\delta)\Big) \\
&\quad = (N!)^n \int_{\prod_{i=1}^n(\Delta_R(\mu_i;N,m,\delta)\cap\mathbb{R}_\le^N)} \frac{1}{(N!)^n} \\
&\qquad\qquad \times \sum_{\sigma_1,\dots,\sigma_n\in S_N} \exp\big(N\kappa_N(h(\sigma_1(\mathbf{x}_1),\dots,\sigma_n(\mathbf{x}_n)))\big) \prod_{i=1}^n d\mathbf{x}_i. \qquad (3.2)
\end{aligned}
$$

Let $\Xi(N) = (\xi_1(N),\dots,\xi_n(N))$ be an approximating sequence for $(\mu_1,\dots,\mu_n)$. For any $\varepsilon > 0$ there exists a real polynomial p of variables $x_1,\dots,x_n$ such that $\|p-h\| < \varepsilon$. Then there exist an $m \in \mathbb{N}$, a $\delta > 0$ and an $N_0 \in \mathbb{N}$ such that, for every $N \ge N_0$, if $\mathbf{x}_i \in \Delta_R(\mu_i;N,m,\delta)\cap\mathbb{R}_\le^N$ for $1 \le i \le n$, then we have

$$
|\kappa_N(p(\sigma_1(\mathbf{x}_1),\dots,\sigma_n(\mathbf{x}_n)) - \kappa_N(p(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N))))| < \varepsilon
$$

so that

$$
|\kappa_N(h(\sigma_1(\mathbf{x}_1),\dots,\sigma_n(\mathbf{x}_n)) - \kappa_N(h(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N))))| < 3\varepsilon
$$

for all $\sigma_1,\dots,\sigma_n \in S_N$. Hence by (3.2) we obtain

$$
\begin{aligned}
Z_h^N &\ge Z_h^N \mu_h^{\otimes N}\Big(\prod_{i=1}^n \Delta_R(\mu_i;N,m,\delta)\Big) \\
&\ge e^{-3N\varepsilon}\frac{1}{(N!)^n} \sum_{\sigma_1,\dots,\sigma_n\in S_N} \exp\big(N\kappa_N(h(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N))))\big) \\
&\qquad \times \prod_{i=1}^n \lambda_N(\Delta_R(\mu_i;N,m,\delta)) \qquad (3.3)
\end{aligned}
$$

and

$$
\begin{aligned}
&Z_h^N \mu_h^{\otimes N}\Big(\prod_{i=1}^n \Delta_R(\mu_i;N,m,\delta)\Big) \\
&\quad \le e^{3N\varepsilon}\frac{1}{(N!)^n} \sum_{\sigma_1,\dots,\sigma_n\in S_N} \exp\big(N\kappa_N(h(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N))))\big) \\
&\qquad \times \prod_{i=1}^n \lambda_N(\Delta_R(\mu_i;N,m,\delta)). \qquad (3.4)
\end{aligned}
$$

It follows from (3.3) that

$$
\begin{aligned}
P(h) &= \frac{1}{N}\log Z_h^N \\
&\ge -3\varepsilon + \frac{1}{N}\log\Big[\frac{1}{(N!)^n} \sum_{\sigma_1,\dots,\sigma_n\in S_N} \exp\big(N\kappa_N(h(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N))))\big)\Big] \\
&\qquad + \sum_{i=1}^n \frac{1}{N}\log\lambda_N(\Delta_R(\mu_i;N,m,\delta)).
\end{aligned}
$$

This yields

$$P(h) \geq -3\varepsilon + P_{\mathrm{sym}}(h : \mu_1, \ldots, \mu_n) + \sum_{i=1}^{n} \lim_{N\to\infty} \frac{1}{N} \log \lambda_N(\Delta_R(\mu_i; N, m, \delta))$$

thanks to the existence of the limits in the last term. Letting $m \to \infty$ and $\delta \searrow 0$ gives

$$P(h) \geq -3\varepsilon + P_{\mathrm{sym}}(h : \mu_1, \ldots, \mu_n) + \sum_{i=1}^{n} H(\mu_i),$$

which implies inequality (3.1) since $\varepsilon > 0$ is arbitrary.

Next let us prove the equivalence of conditions (i)–(iii). For $1 \leq i \leq n$ let $\mu_{h,i}$ be the ith marginal of μ_h. Since

$$\begin{aligned} d\mu_{h,i}(x) &= \frac{1}{Z_h} \biggl(\int_{[-R,R]^{n-1}} e^{h(x_1,\ldots,x_{i-1},x,x_{i+1}\ldots,x_n)}\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n \biggr)\, dx \\ &= \frac{1}{Z_h}\, e^{h_i(x)}\, dx, \end{aligned} \tag{3.5}$$

we notice that $\mu_{h,i}$ is the Gibbs measure associated with h_i for $1 \leq i \leq n$. Hence (ii)$\Leftrightarrow$(iii) follows. To prove (ii) $\Rightarrow$ (i), assume that $\mu_i = \mu_{h,i}$ for all $i = 1, \ldots, n$. Then Lemma 3.1 (a) gives

$$\begin{aligned} &\lim_{N\to\infty} \mu_h^{\otimes N}\bigl(\{(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in ([-R,R]^N)^n : \mathbf{x}_i \in \Delta_R(\mu_i; N, m, \delta)\}\bigr) \\ &\quad = \lim_{N\to\infty} \mu_i^{\otimes N}(\Delta_R(\mu_i; N, m, \delta)) = 1. \end{aligned}$$

Therefore,

$$\begin{aligned} &\lim_{N\to\infty} \mu_h^{\otimes N}\Bigl(\prod_{i=1}^{n} \Delta_R(\mu_i; N, m, \delta)\Bigr) \\ &\qquad = \lim_{N\to\infty} \mu_h^{\otimes N}\Bigl(\bigcap_{i=1}^{n} \{(\mathbf{x}_1, \ldots, \mathbf{x}_n) : \mathbf{x}_i \in \Delta_R(\mu_i; N, m, \delta)\}\Bigr) = 1. \end{aligned}$$

Hence it follows from (3.4) that

$$P(h) \leq 3\varepsilon + P_{\mathrm{sym}}(h : \mu_1, \ldots, \mu_n) + \sum_{i=1}^{n} H(\mu_i),$$

which implies equality in (i).

Conversely, assume (i). Since (3.3) implies that

$$\begin{aligned} P(h) + \limsup_{N\to\infty} \frac{1}{N} \log \mu_h^{\otimes N}\Bigl(\prod_{i=1}^{n} \Delta_R(\mu_i; N, m, \delta)\Bigr) \\ \geq -3\varepsilon + P_{\mathrm{sym}}(h : \mu_1, \ldots, \mu_n) + \sum_{i=1}^{n} H(\mu_i), \end{aligned}$$

we have

$$\limsup_{N\to\infty} \frac{1}{N} \log \mu_h^{\otimes N}\Bigl(\prod_{i=1}^{n} \Delta_R(\mu_i; N, m, \delta)\Bigr) \geq -3\varepsilon.$$

Here we can take m arbitrarily large and $\delta > 0$ arbitrarily small for any given $\varepsilon > 0$. Therefore,

$$\limsup_{N\to\infty} \frac{1}{N} \log \mu_h^{\otimes N}\Big(\prod_{i=1}^n \Delta_R(\mu_i; N, m, \delta)\Big) = 0$$

for all $m \in \mathbb{N}$ and all $\delta > 0$. Since

$$\mu_h^{\otimes N}\Big(\prod_{i=1}^n \Delta_R(\mu_i; N, m, \delta)\Big) \le \mu_{h,i}^{\otimes N}(\Delta_R(\mu_i; N, m, \delta)),$$

we have

$$\limsup_{N\to\infty} \frac{1}{N} \log \mu_{h,i}^{\otimes N}(\Delta_R(\mu_i; N, m, \delta)) = 0$$

for all $m \in \mathbb{N}$, $\delta > 0$ and $i = 1, \dots, n$. Lemma 3.1 (b) implies that $\mu_i = \mu_{h,i}$ for all $i = 1, \dots, n$, so (ii) holds. ∎

REMARK 3.3. Let $\Xi(N) = (\xi_1(N), \dots, \xi_n(N))$ be an approximating sequence for the n-tuple $(\mu_1, \dots, \mu_n)$. Let $h_1, \dots, h_n \in C_{\mathbb{R}}([-R, R])$ and consider h_i as an element of $C_{\mathbb{R}}([-R, R]^n)$ depending on the ith variable x_i, $1 \le i \le n$, so that $(h_1 + \dots + h_n)(\mathbf{x}) = h_1(x_1) + \dots + h_n(x_n)$ for $\mathbf{x} = (x_1, \dots, x_n)$. Since

$$\begin{aligned}
\frac{1}{(N!)^n} &\sum_{\sigma_1,\dots,\sigma_n \in S_N} \exp\big(N\kappa_N((h - (h_1 + \dots + h_n))(\sigma_1(\xi_1(N)), \dots, \sigma_n(\xi_n(N))))\big) \\
&= \frac{1}{(N!)^n} \sum_{\sigma_1,\dots,\sigma_n \in S_N} \exp\big(N\kappa_N(h(\sigma_1(\xi_1(N)), \dots, \sigma_n(\xi_n(N))))\big) \\
&\qquad \times \prod_{i=1}^n \exp\big(-N\kappa_N(h_i(\xi_i(N)))\big)
\end{aligned}$$

and $\lim_{N\to\infty} \kappa_N(h_i(\xi_i(N))) = \mu_i(h_i)$, it follows that

$$\begin{aligned}
P_{\mathrm{sym}}(h - (h_1 + \dots + h_n) : \mu_1, \dots, \mu_n) + \sum_{i=1}^n P(h_i) \\
= P_{\mathrm{sym}}(h : \mu_1, \dots, \mu_n) + \sum_{i=1}^n (-\mu_i(h_i) + P(h_i)).
\end{aligned}$$

Hence we notice that

$$\begin{aligned}
P_{\mathrm{sym}}&(h : \mu_1, \dots, \mu_n) + \sum_{i=1}^n H(\mu_i) \\
&= \inf_{h_1,\dots,h_n} \Big\{ P_{\mathrm{sym}}(h - (h_1 + \dots + h_n) : \mu_1, \dots, \mu_n) + \sum_{i=1}^n P(h_i) \Big\},
\end{aligned}$$

where $h_1 + \dots + h_n$ is given as above for $h_1, \dots, h_n \in C_{\mathbb{R}}([-R, R])$. In particular, when μ_i is the Gibbs measure associated with h_i for $1 \le i \le n$, we have

$$P_{\mathrm{sym}}(h : \mu_1, \dots, \mu_n) + \sum_{i=1}^n H(\mu_i) = P_{\mathrm{sym}}(h - (h_1 + \dots + h_n) : \mu_1, \dots, \mu_n) + \sum_{i=1}^n P(h_i).$$

Hence, if the equivalent conditions (i)–(iii) of Theorem 3.2 are satisfied, then the equality

$$P(h) = P_{\mathrm{sym}}(h - (h_1 + \cdots + h_n) : \mu_1, \ldots, \mu_n) + \sum_{i=1}^{n} P(h_i)$$

holds as well for $h_1, \ldots, h_n$ given in (iii).

The next lemma is concerned with general relation between $\mathcal{I}_{\mathrm{sym}}(\mu)$ and $I_{\mathrm{sym}}(\mu)$.

LEMMA 3.4. *$\mathcal{I}_{\mathrm{sym}}(\mu) \le I_{\mathrm{sym}}(\mu)$ for every $\mu \in \mathrm{Prob}([-R,R]^n)$.*

Proof. Let $\mu \in \mathrm{Prob}([-R,R]^n)$ and $\mu_1, \ldots, \mu_n$ be the marginals of μ, and choose an approximating sequence $\Xi(N) = (\xi_1(N), \ldots, \xi_n(N))$ for $(\mu_1, \ldots, \mu_n)$. It suffices to prove that

$$I_{\mathrm{sym}}(\mu) \ge \mu(p) - P_{\mathrm{sym}}(p : \mu_1, \ldots, \mu_n)$$

for all real polynomials p of variables $x_1, \ldots, x_n$. For any $\varepsilon > 0$ there exist an $m \in \mathbb{N}$ and a $\delta > 0$ such that, for every $N \in \mathbb{N}$, if $(\sigma_1, \ldots, \sigma_n) \in \Delta_{\mathrm{sym}}(\mu : \Xi(N); N, m, \delta)$ then

$$|\kappa_N(p(\sigma_1(\xi_1(N)), \ldots, \sigma_n(\xi_n(N)))) - \mu(p)| < \varepsilon$$

so that

$$e^{N(\mu(p)-\varepsilon)} < \exp\bigl(N\kappa_N(p(\sigma_1(\xi_1(N)), \ldots, \sigma_n(\xi_n(N))))\bigr).$$

Therefore,

$$\begin{aligned} &e^{N(\mu(p)-\varepsilon)} \frac{1}{(N!)^n} \#\Delta_{\mathrm{sym}}(\mu : \Xi(N); N, m, \delta) \\ &\quad \le \frac{1}{(N!)^n} \sum_{(\sigma_1,\ldots,\sigma_n) \in \Delta_{\mathrm{sym}}(\mu:\Xi(N);N,m,\delta)} \exp\bigl(N\kappa_N(p(\sigma_1(\xi_1(N)), \ldots, \sigma_n(\xi_n(N))))\bigr) \\ &\quad \le \frac{1}{(N!)^n} \sum_{\sigma_1,\ldots,\sigma_n \in S_N} \exp\bigl(N\kappa_N(p(\sigma_1(\xi_1(N)), \ldots, \sigma_n(\xi_n(N))))\bigr), \end{aligned}$$

which implies that

$$\mu(p) - \varepsilon - I_{\mathrm{sym}}(\mu) \le P_{\mathrm{sym}}(p : \mu_1, \ldots, \mu_n).$$

This gives the desired inequality since $\varepsilon > 0$ is arbitrary. ■

The next theorem gives an exact relation between $\mathcal{I}_{\mathrm{sym}}(\mu)$ and $H(\mu)$.

THEOREM 3.5. *For every $\mu \in \mathrm{Prob}([-R,R]^n)$ with marginals $\mu_1, \ldots, \mu_n \in \mathrm{Prob}([-R,R])$,*

$$H(\mu) = -\mathcal{I}_{\mathrm{sym}}(\mu) + \sum_{i=1}^{n} H(\mu_i).$$

Moreover, if $H(\mu_i) > -\infty$ for all $i = 1, \ldots, n$, then

$$\mathcal{I}_{\mathrm{sym}}(\mu) = I_{\mathrm{sym}}(\mu) = S(\mu, \mu_1 \otimes \cdots \otimes \mu_n),$$

and $\mathcal{I}_{\mathrm{sym}}(\mu) = 0$ if and only if $\mu = \mu_1 \otimes \cdots \otimes \mu_n$, i.e., the coordinate variables $x_1, \ldots, x_n$ are independent with respect to μ.

Proof. By (3.1) and Definition 2.4, for every $h \in C_{\mathbb{R}}([-R,R]^n)$ we have

$$\begin{aligned} -\mu(h) + P(h) &\geq -\mu(h) + P_{\mathrm{sym}}(h : \mu_1, \dots, \mu_n) + \sum_{i=1}^n H(\mu_i) \\ &\geq -\mathcal{I}_{\mathrm{sym}}(\mu) + \sum_{i=1}^n H(\mu_i). \end{aligned} \tag{3.6}$$

Hence by (1.1), Lemma 3.4 and Theorem 1.2 we have

$$H(\mu) \geq -\mathcal{I}_{\mathrm{sym}}(\mu) + \sum_{i=1}^n H(\mu_i) \geq -I_{\mathrm{sym}}(\mu) + \sum_{i=1}^n H(\mu_i) = H(\mu)$$

so that the first assertion is proved. The second assertion immediately follows from the first and [5, Corollary 1.7]. ■

PROPOSITION 3.6. *Let $h \in C_{\mathbb{R}}([-R,R]^n)$ and $\mu \in \mathrm{Prob}([-R,R]^n)$. Let $\mu_1, \dots, \mu_n$ be the marginals of μ and $h_1, \dots, h_n$ be as given in (iii) of Theorem 3.2. Then the following are equivalent:*

(i) μ is the Gibbs measure associated with h;
(ii) μ is mutually equilibrium associated with h and μ_i is the Gibbs measure associated with h_i for each $i = 1, \dots, n$.

Proof. (i) ⇒ (ii). Assume that μ is the Gibbs measure associated with h. By (3.6) and Theorem 3.5,

$$\begin{aligned} H(\mu) = -\mu(h) + P(h) &\geq -\mu(h) + P_{\mathrm{sym}}(h : \mu_1, \dots, \mu_n) + \sum_{i=1}^n H(\mu_i) \\ &\geq -\mathcal{I}_{\mathrm{sym}}(\mu) + \sum_{i=1}^n H(\mu_i) = H(\mu). \end{aligned}$$

Moreover, since μ_i is the ith marginal of $\mu = \mu_h$, it follows as in the proof of Theorem 3.2 (see (3.5)) that μ_i is the Gibbs measure associated with h_i for $1 \leq i \leq n$. In particular, $H(\mu_i) > -\infty$ for all $i = 1, \dots, n$. Hence

$$-\mathcal{I}_{\mathrm{sym}}(\mu) = -\mu(h) + P_{\mathrm{sym}}(h : \mu_1, \dots, \mu_n),$$

that is, μ is mutually equilibrium associated with h.

(ii) ⇒ (i). Assume (ii). By Theorems 3.5 and 3.2,

$$\begin{aligned} H(\mu) &= -\mathcal{I}_{\mathrm{sym}}(\mu) + \sum_{i=1}^n H(\mu_i) = -\mu(h) + P_{\mathrm{sym}}(h : \mu_1, \dots, \mu_n) + \sum_{i=1}^n H(\mu_i) \\ &= -\mu(h) + P(h) \end{aligned}$$

so that (i) follows. ■

4. The discrete case. In information theory, random variables mostly take values in a discrete set of alphabets and the basic quantity is the Shannon entropy rather than the Boltzmann-Gibbs entropy. So the discrete versions of the preceding results in Sections 2 and 3 are of even more importance; they are presented in this section.

Let $\mathcal{X} = \{t_1, \dots, t_d\}$ be a finite set of alphabets and consider the n-fold product $\mathcal{X}^n$. The *Shannon entropy* of a probability measure $\mu \in \mathrm{Prob}(\mathcal{X})$ is

$$S(\mu) := -\sum_{t \in \mathcal{X}} \mu(t) \log \mu(t).$$

For each sequence $\mathbf{x} = (x_1, \dots, x_N) \in \mathcal{X}^N$, the *type* of $\mathbf{x}$ is a probability measure on $\mathcal{X}$ given by

$$\nu_{\mathbf{x}}(t) := \frac{N_{\mathbf{x}}(t)}{N} \quad \text{where} \quad N_{\mathbf{x}}(t) := \#\{j : x_j = t\}, \ \ t \in \mathcal{X}.$$

For each $\mu \in \mathrm{Prob}(\mathcal{X})$ (resp. $\mu \in \mathrm{Prob}(\mathcal{X}^n)$) and for each $N \in \mathbb{N}$ and $\delta > 0$ we denote by $\Delta(\mu; N, \delta)$ the set of all sequences $\mathbf{x} \in \mathcal{X}^N$ (resp. $\mathbf{x} \in (\mathcal{X}^n)^N$) such that $|\nu_{\mathbf{x}}(t) - \mu(t)| < \delta$ for all $t \in \mathcal{X}$ (resp. $t \in \mathcal{X}^n$), that is, $\Delta(\mu; N, \delta)$ is the set of all δ-typical sequences (with respect to μ). The Shannon entropy has the following limiting formula:

$$S(\mu) = \lim_{\delta \searrow 0} \lim_{N \to \infty} \frac{1}{N} \log \#\Delta(\mu; N, \delta) \tag{4.1}$$

(see [1, 2] and also [5, §2] for a concise exposition).

For $N \in \mathbb{N}$ let $\mathcal{X}^N_{\leq}$ denote the set of all sequences of length N of the form

$$\mathbf{x} = (t_1, \dots, t_1, t_2, \dots, t_2, \dots, t_d, \dots, t_d)$$

so that $\mathcal{X}^N_{\leq}$ is regarded as the set of all types from $\mathcal{X}^N$. The action of S_N on $\mathcal{X}^N$ is similar to that on $\mathbb{R}^N$ given in Definition 1.1.

Definition 4.1. Let $\mu \in \mathrm{Prob}(\mathcal{X}^n)$ and $\mu_i \in \mathrm{Prob}(\mathcal{X})$ be the ith marginal of μ for $1 \leq i \leq n$. Choose an approximating sequence $\Xi(N) = (\xi_1(N), \dots, \xi_n(N))$, $N \in \mathbb{N}$, for $(\mu_1, \dots, \mu_n)$, that is, $\xi_i(N) \in \mathcal{X}^N_{\leq}$ and $\nu_{\xi_i(N)}(t) \to \mu_i(t)$ as $N \to \infty$ for all $t \in \mathcal{X}$ and $i = 1, \dots, n$. For each $N \in \mathbb{N}$ and $\delta > 0$ we define $\Delta_{\mathrm{sym}}(\mu : \Xi(N); N, \delta)$ to be the set of all $(\sigma_1, \dots, \sigma_n) \in S_N^n$ such that

$$(\sigma_1(\xi_1(N)), \dots, \sigma_n(\xi_n(N))) \in \Delta(\mu; N, \delta).$$

We define

$$I_{\mathrm{sym}}(\mu) := -\lim_{\delta \searrow 0} \limsup_{N \to \infty} \frac{1}{N} \log \gamma_{S_N}^{\otimes n}(\Delta_{\mathrm{sym}}(\mu : \Xi(N); N, \delta))$$

and $\overline{I}_{\mathrm{sym}}(\mu)$ by replacing lim sup by lim inf. See [5, Lemma 2.4] for the independence of the choice of $\Xi(N)$ for $I_{\mathrm{sym}}(\mu)$ and $\overline{I}_{\mathrm{sym}}(\mu)$ as well as their equivalent definitions.

The two quantities $I_{\mathrm{sym}}(\mu)$ and $\overline{I}_{\mathrm{sym}}(\mu)$ are equal and connected to $S(\mu)$ as follows.

Theorem 4.2 ([5, Theorem 2.5]). *For every $\mu \in \mathrm{Prob}(\mathcal{X}^n)$ with marginals $\mu_1, \dots, \mu_n \in \mathrm{Prob}(\mathcal{X})$,*

$$I_{\mathrm{sym}}(\mu) = \overline{I}_{\mathrm{sym}}(\mu) = -S(\mu) + \sum_{i=1}^{n} S(\mu_i).$$

We denote by $C_{\mathbb{R}}(\mathcal{X}^n)$ the real Banach space of real functions on $\mathcal{X}^n$ with the norm $\|f\| := \max\{|f(\mathbf{x})| : \mathbf{x} \in \mathcal{X}^n\}$.

DEFINITION 4.3. Let $\mu_1, \ldots, \mu_n \in \mathrm{Prob}(\mathcal{X})$ and choose an approximating sequence $\Xi(N) = (\xi_1(N), \ldots, \xi_n(N))$ for $(\mu_1, \ldots, \mu_n)$ as given in Definition 4.1. For each $h \in C_{\mathbb{R}}(\mathcal{X}^n)$ and $\mathbf{x}_i \in \mathcal{X}^N$, $1 \le i \le n$, define $h(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $\kappa_N(h(\mathbf{x}_1, \ldots, \mathbf{x}_n))$ in the same manner as in (2.1) and (2.2) so that

$$\kappa_N(h(\mathbf{x}_1, \ldots, \mathbf{x}_n)) = \sum_{t \in \mathcal{X}^n} h(t)\nu_{(\mathbf{x}_1, \ldots, \mathbf{x}_n)}(t)$$

for $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ regarded as a sequence in $(\mathcal{X}^n)^N$. We define the *mutual pressure* of h with respect to $(\mu_1, \ldots, \mu_n)$ to be

$$\begin{aligned} &P_{\mathrm{sym}}(h : \mu_1, \ldots, \mu_n) \\ &\quad := \limsup_{N \to \infty} \frac{1}{N} \log\Bigg[\frac{1}{(N!)^n} \sum_{\sigma_1, \ldots, \sigma_n \in S_N} \exp\big(N\kappa_N(h(\sigma_1(\xi_1(N)), \ldots, \sigma_n(\xi_n(N))))\big)\Bigg]. \end{aligned}$$

Moreover, for each $\mu \in \mathrm{Prob}(\mathcal{X}^n)$ with marginals $\mu_1, \ldots, \mu_n \in \mathrm{Prob}(\mathcal{X})$ we define

$$\mathcal{I}_{\mathrm{sym}}(\mu) := \sup\{\mu(h) - P_{\mathrm{sym}}(h : \mu_1, \ldots, \mu_n) : h \in C_{\mathbb{R}}(\mathcal{X}^n)\},$$

and we say that μ is *mutually equilibrium* associated with h if the equality

$$\mathcal{I}_{\mathrm{sym}}(\mu) = \mu(h) - P_{\mathrm{sym}}(h : \mu_1, \ldots, \mu_n)$$

holds.

Then all the results in Section 2 are valid in this discrete setting as well. To see this, it is convenient to reduce the discrete case to a special case of the continuous case of Section 2 in the following way. Choose d points $\hat{t}_1 < \hat{t}_2 < \cdots < \hat{t}_d$ in $[-R, R]$ corresponding to $t_1, t_2, \ldots, t_d$ in $\mathcal{X}$. For each $\mu \in \mathrm{Prob}(\mathcal{X}^n)$ with marginals $\mu_1, \ldots, \mu_n$ we have the corresponding (atomic) probability measure $\hat{\mu} \in \mathrm{Prob}([-R, R]^n)$ given by

$$\hat{\mu} := \sum_{\mathbf{x} \in \mathcal{X}^n} \mu(\mathbf{x})\delta_{\hat{\mathbf{x}}},$$

and similarly $\hat{\mu}_1, \ldots, \hat{\mu}_n \in \mathrm{Prob}([-R, R])$, $1 \le i \le n$. Then the marginals of $\hat{\mu}$ are $\hat{\mu}_1, \ldots, \hat{\mu}_n$. For each approximating sequence $(\xi_1(N), \ldots, \xi_n(N))$ for $(\mu_1, \ldots, \mu_n)$ we have the corresponding $\hat{\xi}_i(N) \in [-R, R]^N_{\le}$, $1 \le i \le n$. Since for all $k \in \mathbb{N}$,

$$\kappa_N(\hat{\xi}_i(N)^k) = \sum_{t \in \mathcal{X}} \hat{t}^k \nu_{\xi_i(N)}(t) \to \sum_{t \in \mathcal{X}} \hat{t}^k \mu_i(t) = \int x^k \, d\hat{\mu}_i(x) \quad \text{as } N \to \infty,$$

it follows that $(\hat{\xi}_1(N), \ldots, \hat{\xi}_n(N))$ is an approximating sequence for $(\hat{\mu}_1, \ldots, \hat{\mu}_n)$. For each $h \in C_{\mathbb{R}}(\mathcal{X}^n)$ choose an $\hat{h} \in C_{\mathbb{R}}([-R, R]^n)$ such that $\hat{h}(\hat{\mathbf{x}}) = h(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}^n$. Then we notice that $P_{\mathrm{sym}}(h : \mu_1, \ldots, \mu_n)$ in Definition 4.3 is equal to $P_{\mathrm{sym}}(\hat{h} : \hat{\mu}_1, \ldots, \hat{\mu}_n)$ defined in Definition 2.1, and that $\mathcal{I}_{\mathrm{sym}}(\mu)$ in Definition 4.3 is equal to $\mathcal{I}_{\mathrm{sym}}(\hat{\mu})$ defined in Definition 2.4. Upon these considerations it is rather straightforward to show the discrete versions of the results in Section 2. For example, for $h, h' \in C_{\mathbb{R}}(\mathcal{X}^n)$ choose $\hat{h}, \hat{g} \in C_{\mathbb{R}}([-R, R]^n)$ such that $\hat{h}|_{\mathcal{X}^n} = h$, $\|\hat{h}\| = \|h\|$, $\hat{g}|_{\mathcal{X}^n} = h - h'$ and $\|\hat{g}\| = \|h - h'\|$, and define $\hat{h}' := \hat{h} - \hat{g}$. Then $\hat{h}'|_{\mathcal{X}^n} = h'$ and $\|\hat{h} - \hat{h}'\| = \|h - h'\|$. Hence the discrete version of Proposition 2.3 (3) is seen as follows:

$$\begin{aligned}&|P_{\rm sym}(h:\mu_1,\dots,\mu_n)-P_{\rm sym}(h':\mu_1,\dots,\mu_n)|\\&\qquad=|P_{\rm sym}(\hat h:\hat\mu_1,\dots,\hat\mu_n)-P_{\rm sym}(\hat h':\hat\mu_1,\dots,\hat\mu_n)|\le\|\hat h-\hat h'\|=\|h-h'\|.\end{aligned}$$

Also, the discrete version of Proposition 2.5 is seen as follows:

$$\begin{aligned}P_{\rm sym}(h:\mu_1,\dots,\mu_n)&=P_{\rm sym}(\hat h:\hat\mu_1,\dots,\hat\mu_n)\\&=\max\{\lambda(\hat h)-\mathcal{I}_{\rm sym}(\lambda):\lambda\in{\rm Prob}_{\hat\mu_1,\dots,\hat\mu_n}([-R,R]^n)\}\\&=\max\{\mu(h)-\mathcal{I}_{\rm sym}(\mu):\mu\in{\rm Prob}_{\mu_1,\dots,\mu_n}(\mathcal{X}^n)\}\end{aligned}$$

since ${\rm Prob}_{\hat\mu_1,\dots,\hat\mu_n}([-R,R]^n)=\{\hat\mu:\mu\in{\rm Prob}_{\mu_1,\dots,\mu_n}(\mathcal{X}^n)\}$.

Now let us show the discrete version of Theorem 3.2. Although the proof is essentially the same as that of Theorem 3.2, some non-trivial modifications are necessary due to the difference between the Shannon and Boltzmann-Gibbs entropies.

THEOREM 4.4. *For every $h\in C_{\mathbb{R}}(\mathcal{X}^n)$ and every $\mu_1,\dots,\mu_n\in{\rm Prob}(\mathcal{X})$,*

$$P(h)\ge P_{\rm sym}(h:\mu_1,\dots,\mu_n)+\sum_{i=1}^n S(\mu_i),\tag{4.2}$$

and the following conditions are equivalent:

(i) $P(h)=P_{\rm sym}(h:\mu_1,\dots,\mu_n)+\sum_{i=1}^n S(\mu_i)$*;*

(ii) $\mu_1,\dots,\mu_n$ *are the marginals of the Gibbs measure μ_h associated with h given by*

$$\mu_h(\mathbf{x}):=\frac{1}{Z_h}\,e^{h(\mathbf{x})},\quad \mathbf{x}\in\mathcal{X}^n\quad\textit{with}\quad Z_h:=\sum_{\mathbf{x}\in\mathcal{X}^n}e^{h(\mathbf{x})};\tag{4.3}$$

(iii) for each $i=1,\dots,n$, μ_i is the Gibbs measure associated with $h_i\in C_{\mathbb{R}}(\mathcal{X})$ defined by

$$h_i(x):=\log\sum_{x_1,\dots,x_{i-1},x_{i+1},\dots,x_n\in\mathcal{X}}e^{h(x_1,\dots,x_{i-1},x,x_{i+1},\dots,x_n)}\qquad\textit{for } x\in\mathcal{X}.$$

Proof. Let μ_h be the Gibbs measure given in (4.3), and let $(\xi_1(N),\dots,\xi_n(N))$ be an approximating sequence for $(\mu_1,\dots,\mu_n)$. For any $\varepsilon>0$ one can choose a $\delta>0$ such that for every $i=1,\dots,n$ and every $p\in{\rm Prob}(\mathcal{X})$, if $|p(t)-\mu_i(t)|<\delta$ for all $t\in\mathcal{X}$, then $|S(p)-S(\mu_i)|<\varepsilon/n$. This means that for each $N\in\mathbb{N}$ and $i=1,\dots,n$, one has $|S(\nu_{\mathbf{x}})-S(\mu_i)|<\varepsilon/n$ whenever $\mathbf{x}\in\Delta(\mu_i;N,\delta)$. Furthermore, when $\delta>0$ is small enough, one can find an $N_0\in\mathbb{N}$ such that, for every $N\ge N_0$, if $\mathbf{x}_i\in\Delta(\mu_i;N,\delta)\cap\mathcal{X}_\le^N$ for $1\le i\le n$, then

$$|\kappa_N(h(\sigma_1(\mathbf{x}_1),\dots,\sigma_n(\mathbf{x}_n))-\kappa_N(h(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N))))|<\varepsilon\tag{4.4}$$

for all $(\sigma_1,\dots,\sigma_n)\in S_N$.

For each sequence $(N_1,\dots,N_d)$ of integers $N_l\ge0$ with $\sum_{l=1}^d N_l=N$, let $S(N_1,\dots,N_d)$ denote the subgroups of S_N consisting of products of permutations of $\{1,\dots,N_1\}$, $\{N_1+1,\dots,N_1+N_2\}$, …, $\{N_1+\cdots+N_{d-1}+1,\dots,N\}$, and let $S_N/S(N_1,\dots,N_d)$ be the set of left cosets of $S(N_1,\dots,N_d)$. For each $\mathbf{x}\in\mathcal{X}_\le^N$ we write $S_{\mathbf{x}}$ for $S(N_{\mathbf{x}}(t_1),\dots,N_{\mathbf{x}}(t_d))$.

For $N \in \mathbb{N}$ it then follows that

$$
\begin{aligned}
&Z_h^N \mu_h^{\otimes n}\Bigl(\prod_{i=1}^n \Delta(\mu_i; N, \delta)\Bigr) \\
&\quad = \sum_{(\mathbf{x}_1,\dots,\mathbf{x}_n)\in\prod_{i=1}^n \Delta(\mu_i;N,\delta)} \exp\bigl(N\kappa_N(h(\mathbf{x}_1,\dots,\mathbf{x}_n))\bigr) \\
&\quad = \sum_{(\mathbf{x}_1,\dots,\mathbf{x}_n)\in\prod_{i=1}^n \bigl(\Delta(\mu_i;N,\delta)\cap\mathcal{X}_{\le}^N\bigr)} \\
&\qquad\qquad \sum_{([\sigma_1],\dots,[\sigma_n])\in(S_N/S_{\mathbf{x}_1},\dots,S_N/S_{\mathbf{x}_n})} \exp\bigl(N\kappa_N(h(\sigma_1(\mathbf{x}_1),\dots,\sigma_n(\mathbf{x}_n)))\bigr),
\end{aligned}
\tag{4.5}
$$

where $\prod_{i=1}^n \Delta(\mu_i; N, \delta)$ in the left-hand side is regarded as a subset of $(\mathcal{X}^n)^N$ in the same manner as in the beginning of the proof of Theorem 3.2, and $[\sigma_i]$ denotes the coset of $S_{\mathbf{x}_i}$ containing σ_i. Moreover we have

$$
\begin{aligned}
&\sum_{\sigma_1,\dots,\sigma_n\in S_N} \exp\bigl(N\kappa_N(h(\sigma_1(\mathbf{x}_1),\dots,\sigma_n(\mathbf{x}_n)))\bigr) \\
&= \sum_{([\sigma_1],\dots,[\sigma_n])\in(S_N/S_{\mathbf{x}_1},\dots,S_N/S_{\mathbf{x}_n})} \Bigl(\prod_{i=1}^n\prod_{l=1}^d N_{\mathbf{x}_i}(t_l)!\Bigr) \exp\bigl(N\kappa_N(h(\sigma_1(\mathbf{x}_1),\dots,\sigma_n(\mathbf{x}_n)))\bigr).
\end{aligned}
\tag{4.6}
$$

For each $i = 1,\dots,n$ and for any $\mathbf{x} \in \mathcal{X}^N$, the Stirling formula implies that

$$
\begin{aligned}
&\frac{1}{N}\sum_{l=1}^d \log N_{\mathbf{x}}(t_l)! - \frac{1}{N}\log N! \\
&\quad = \frac{1}{N}\sum_{l=1}^d \Bigl(N_{\mathbf{x}}(t_l)\log N_{\mathbf{x}}(t_l) - N_{\mathbf{x}}(t_l) + \frac{1}{2}\log N_{\mathbf{x}}(t_l) + O(1)\Bigr) \\
&\qquad - \frac{1}{N}\Bigl(N\log N - N + \frac{1}{2}\log N + O(1)\Bigr) \\
&\quad = \sum_{l=1}^d \frac{N_{\mathbf{x}}(t_l)}{N}\log N_{\mathbf{x}}(t_l) - \log N + o(1) \\
&\quad = -S(\nu_{\mathbf{x}}) + o(1) \quad \text{as } N\to\infty,
\end{aligned}
$$

where $o(1)$ as $N \to \infty$ is uniform for $\mathbf{x} \in \mathcal{X}^N$. Thanks to the above choice of $\delta > 0$, for every $(\mathbf{x}_1,\dots,\mathbf{x}_n) \in \prod_{i=1}^n \Delta(\mu_i; N, \delta)$ we have

$$
\begin{aligned}
&\exp\Bigl[N\Bigl(-\sum_{i=1}^n S(\mu_i) - \varepsilon + o(1)\Bigr)\Bigr] \\
&\quad \le \frac{\prod_{i=1}^n\prod_{l=1}^d N_{\mathbf{x}_i}(t_i)}{(N!)^n} \le \exp\Bigl[N\Bigl(-\sum_{i=1}^n S(\mu_i) + \varepsilon + o(1)\Bigr)\Bigr] \quad \text{as } N\to\infty,
\end{aligned}
\tag{4.7}
$$

where $o(1)$ is uniform for $(\mathbf{x}_1,\dots,\mathbf{x}_n) \in \prod_{i=1}^n \Delta(\mu_i; N, \delta)$.

Combining (4.5)–(4.7) yields

$$
\begin{aligned}
&Z_h^N \mu_h^{\otimes N}\Bigl(\prod_{i=1}^n \Delta(\mu_i;N,\delta)\Bigr) \\
&\quad\ge \sum_{(\mathbf{x}_1,\dots,\mathbf{x}_n)\in\prod_{i=1}^n\left(\Delta(\mu_i;N,\delta)\cap\mathcal{X}_\le^N\right)} \frac{1}{(N!)^n} \sum_{\sigma_1,\dots,\sigma_n\in S_N} \exp\bigl(N\kappa_N(h(\sigma_1(\mathbf{x}_1),\dots,\sigma_n(\mathbf{x}_n)))\bigr) \\
&\qquad\qquad \times \exp\Bigl[N\Bigl(\sum_{i=1}^n S(\mu_i)-\varepsilon+o(1)\Bigr)\Bigr]
\end{aligned}
$$

and the reverse inequality with $+\varepsilon$ in place of $-\varepsilon$ in the last term. From this together with (4.4) we obtain

$$
\begin{aligned}
Z_n^N &\ge Z_h^N \mu_h^{\otimes N}\Bigl(\prod_{i=1}^n \Delta(\mu_i;N,\delta)\Bigr) \\
&\ge e^{-2N\varepsilon}\frac{1}{(N!)^n}\sum_{\sigma_1\dots,\sigma_n\in S_N}\exp\bigl(N\kappa_N(h(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N))))\bigr) \\
&\qquad \times\prod_{i=1}^n \#\bigl(\Delta(\mu_i;N,\delta)\cap\mathcal{X}_\le^N\bigr)\cdot\exp\Bigl[N\Bigl(\sum_{i=1}^n S(\mu_i)+o(1)\Bigr)\Bigr] \qquad (4.8)
\end{aligned}
$$

and

$$
\begin{aligned}
&Z_h^N \mu_h^{\otimes N}\Bigl(\prod_{i=1}^n \Delta(\mu_i;N,\delta)\Bigr) \\
&\quad\le e^{2N\varepsilon}\frac{1}{(N!)^n}\sum_{\sigma_1\dots,\sigma_n\in S_N}\exp\bigl(N\kappa_N(h(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N))))\bigr) \\
&\qquad \times\prod_{i=1}^n \#\bigl(\Delta(\mu_i;N,\delta)\cap\mathcal{X}_\le^N\bigr)\cdot\exp\Bigl[N\Bigl(\sum_{i=1}^n S(\mu_i)+o(1)\Bigr)\Bigr] \qquad (4.9)
\end{aligned}
$$

for all $N \ge N_0$. Furthermore, since

$$
\Delta(\mu_i;N,\delta)=\bigl\{\sigma(\mathbf{x}) : \mathbf{x}\in\Delta(\mu_i;N,\delta)\cap\mathcal{X}_\le^N,\ [\sigma]\in S_N/S_{\mathbf{x}}\bigr\}
$$

so that

$$
\#\Delta(\mu_i;N,\delta)=\sum_{\mathbf{x}\in\Delta(\mu_i;N,\delta)\cap\mathcal{X}_\le^N}\frac{N!}{\prod_{l=1}^d N_{\mathbf{x}}(t_l)!},
$$

we have as inequalities in (4.7)

$$
\begin{aligned}
&\#\bigl(\Delta(\mu_i;N,\delta)\cap\mathcal{X}_\le^N\bigr)\cdot\exp\Bigl[N\Bigl(S(\mu_i)-\frac{\varepsilon}{n}+o(1)\Bigr)\Bigr] \\
&\quad\le \#\Delta(\mu_i;N,\delta)\le\#\bigl(\Delta(\mu_i;N,\delta)\cap\mathcal{X}_\le^N\bigr)\cdot\exp\Bigl[N\Bigl(S(\mu_i)+\frac{\varepsilon}{n}+o(1)\Bigr)\Bigr] \quad \text{as } N\to\infty.
\end{aligned}
$$

This and (4.1) imply that

$$-\frac{\varepsilon}{n} \le \liminf_{N\to\infty} \frac{1}{N} \log \#\bigl(\Delta(\mu_i; N, \delta) \cap \mathcal{X}_{\le}^N\bigr) \\ \le \limsup_{N\to\infty} \frac{1}{N} \log \#\bigl(\Delta(\mu_i; N, \delta) \cap \mathcal{X}_{\le}^N\bigr) \le \frac{\varepsilon}{n}. \tag{4.10}$$

It follows from (4.8) that

$$P(h) = \frac{1}{N} \log Z_h^N \\ \ge -2\varepsilon + \frac{1}{N} \log\biggl[\frac{1}{(N!)^n} \sum_{\sigma_1,\dots,\sigma_n \in S_N} \exp\bigl(N\kappa_N(h(\sigma_1(\xi_1(N)),\dots,\sigma_n(\xi_n(N))))\bigr)\biggr] \\ + \sum_{i=1}^n \frac{1}{N} \log \#\bigl(\Delta(\mu_i; N, \delta) \cap \mathcal{X}_{\le}^N\bigr) + \sum_{i=1}^n S(\mu_i) + o(1) \quad \text{as } N \to \infty,$$

which implies that

$$P(h) \ge -3\varepsilon + P_{\mathrm{sym}}(h : \mu_1, \dots, \mu_n) + \sum_{i=1}^n S(\mu_i)$$

thanks to (4.10). Hence inequality (4.2) follows since $\varepsilon > 0$ is arbitrary.

To prove the equivalence of (i)–(iii), let $\mu_{h,i}$ be the ith marginal of μ_h. Then it follows that $\mu_{h,i}(x) = Z_h^{-1} e^{h_i(x)}$ and so $\mu_{h,i}$ is the Gibbs measure associated with h_i for $1 \le i \le n$. Hence (ii)$\Leftrightarrow$(iii) follows. Assume (ii), i.e., that $\mu_i = \mu_{h,i}$ for all $i = 1, \dots, n$. Since we have $\lim_{N\to\infty} \mu_i^{\otimes N}(\Delta(\mu_i; N, \delta)) = 1$ based on the Sanov theorem as in Lemma 3.1 (a), it follows that

$$\lim_{N\to\infty} \mu_h^{\otimes N}\Bigl(\prod_{i=1}^n \Delta(\mu_i; N, \delta)\Bigr) = 1$$

as in the proof of (ii)$\Rightarrow$(i) of Theorem 3.2. Combining this with (4.9) and (4.10) yields

$$P(h) \le 3\varepsilon + P_{\mathrm{sym}}(h : \mu_1, \dots, \mu_n) + \sum_{i=1}^n S(\mu_i),$$

which implies equality in (i). Conversely, assume (i). Then (4.8) and (4.10) imply that

$$\limsup_{N\to\infty} \frac{1}{N} \log \mu_h^{\otimes N}\Bigl(\prod_{i=1}^n \Delta(\mu_i; N, \delta)\Bigr) \ge -3\varepsilon.$$

The same reasoning as in the last part of the proof of Theorem 3.2 gives

$$\limsup_{N\to\infty} \frac{1}{N} \log \mu_{h,i}^{\otimes N}(\Delta(\mu_i; N, \delta)) = 0$$

for all $\delta > 0$ and $i = 1, \dots, n$. Since we have a result similar to Lemma 3.1 (b) in the present discrete situation, it follows that $\mu_i = \mu_{h,i}$ for all $i = 1, \dots, n$, and so (ii) holds. ∎

The next theorem and proposition are the discrete versions of Theorem 3.5 and Proposition 3.6. Since their proofs based on Theorems 4.2 and 4.4 are similar to those in Section 3, we omit the details. Here note only that $\mathcal{I}_{\mathrm{sym}}(\mu) \le I_{\mathrm{sym}}(\mu)$ for every $\mu \in \mathrm{Prob}(\mathcal{X}^n)$ can be shown similarly to the proof of Lemma 3.4 or by the same reasoning as given after

Definition 4.3, and that the Legendre transform expression as in (1.1)

$$S(\mu) = \inf\{-\mu(h) + P(h) : h \in C_{\mathbb{R}}(\mathcal{X}^n)\}$$

is valid for every $\mu \in \mathrm{Prob}(\mathcal{X}^n)$.

THEOREM 4.5. *For every $\mu \in \mathrm{Prob}(\mathcal{X}^n)$ with marginals $\mu_1, \dots, \mu_n \in \mathrm{Prob}(\mathcal{X})$,*

$$\mathcal{I}_{\mathrm{sym}}(\mu) = I_{\mathrm{sym}}(\mu) = -S(\mu) + \sum_{i=1}^{n} S(\mu_i).$$

PROPOSITION 4.6. *Let $h \in C_{\mathbb{R}}(\mathcal{X}^n)$ and $\mu \in \mathrm{Prob}(\mathcal{X}^n)$. Let $\mu_1, \dots, \mu_n$ be the marginals of μ and $h_1, \dots, h_n$ be as given in (iii) of Theorem 4.4. Then the following are equivalent:*

(i) μ is Gibbs measure associated with h;

(ii) μ is mutually equilibrium associated with h and μ_i is the Gibbs measure associated with h_i for each $i = 1, \dots, n$.

Acknowledgments. Research of F. Hiai was supported in part by Grant-in-Aid for Scientific Research (B)17340043.

References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley-Interscience, Hoboken, NJ, 2006.

[2] I. Csiszár and P. C. Shields, *Information theory and statistics: a tutorial*, in: Foundations and Trends in Communications and Information Theory, Vol. 1, No. 4 (2004), 417–528, Now Publishers.

[3] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, 2nd ed., Applications of Mathematics 38, Springer-Verlag, New York, 1998.

[4] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*, Mathematical Surveys and Monographs 77, Amer. Math. Soc., Providence, 2000.

[5] F. Hiai and D. Petz, *A new approach to mutual information*, in: Noncommutative Harmonic Analysis with Applications to Probability, M. Bożejko et al. (eds.), Banach Center Publications 78, 2007, 151–164.

[6] D. Voiculescu, *The analogues of entropy and of Fisher's information measure in free probability theory, II*, Invent. Math. 118 (1994), 411–440.